7N-82-CR
280693

A New Paradigm for Science

Peter J. Denning

September 28, 1987

RIACS Technical Report TR-87.24

NASA Cooperative Agreement Number NCC 2-387

(NASA-CR-181433) A NEW PARADIGM FOR SCIENCE (Research Inst. for Advanced Computer Science) 13 p

N91-70394

Unclas

00/82 0280693

RIACS

Research Institute for Advanced Computer Science

A New Paradigm for Science

Peter J. Denning

September 28, 1987

RIACS Technical Report TR-87.24

NASA Cooperative Agreement Number NCC 2-387

A New Paradigm for Science

Peter J. Denning

Research Institute for Advanced Computer Science
NASA Ames Research Center

RIACS Technical Report TR-87.24
September 28, 1987

Science comprises systematic bodies of information, methods of inquiry and of evaluating hypotheses those inquiries generate, and communities of people working together in each discipline. The knowledge base of science embodies all the information, methods, and know-how of science. Part of it is recorded and part of it lives in the people of science. The science knowledge base has become so large that it is becoming increasingly fragmented and disorganized. To cope with this "information explosion," a new paradigm will emerge for the conduct of science; it will explicitly recognizing living knowledge and the importance of conversations among people, and will be based in networking technology.

This is a preprint of the column *The Science of Computing* for *American Scientist* 75, No 6 (November-December 1987).

Work reported herein was supported in part by Cooperative Agreement NCC 2-387 between the National Aeronautics and Space Administration (NASA) and the Universities Space Research Association (USRA).

A New Paradigm for Science

Peter J. Denning

Research Institute for Advanced Computer Science

September 28, 1987

Most of us think of science as a process for making inquiries about phenomena and for evaluating hypotheses generated by those inquiries. The results are organized bodies of knowledge in various disciplines. Besides its experimental side, the scientific process has an analytic side, which seeks to develop mathematical models of phenomena, and a computational side, which uses the models to make predictions for later experimental validation.

But science is not only the knowledge in various disciplines and the processes for producing that knowledge; it is groups of people working together in each discipline. They are the institutional memory, the selection mechanism, deciding what science is and is not. The knowledge of science and the expertise of science live in the people of science.

This larger view brings into focus an entity I will call the knowledge base of science, which is the sum total of all knowledge and know-how of all the disciplines. Here the term knowledge means something more dynamic than the organized body of facts and data about phenomena; it means "living information," the ability of people to draw inferences from facts and data and to invent new hypotheses. Similarly, the term know-how refers to more than processes and methods; it involves skills that can be developed only through experience and training. Only the facts, data, and algorithms of science can be recorded in books and other media; the rest of the knowledge and know-how lives in the people of science and in their conversations with one another. It is important to notice the distinction between the recorded part of the knowledge base and the newly created, living part.

How large is the science knowledge base? Its enormity can be estimated from the number of working scientists, the total funds expended on R&D, and the number of scholarly publications. According to the National Science Foundation, there are about 825,000 R&D scientists and in the United States *(1)*. About 300,000 persons each year receive bachelor's or other first professional degrees in science and engineering. The corresponding figures worldwide are two or these times larger. There is an element of truth in the quip that half the scientists who ever lived are alive today.

As for expenditures, requests for R&D in the U.S. government's budget for 1988 amounts to almost $60 billion, and when industry's R&D is included, the total expenditure is approximately $120 billion. The competition for these funds has been keen, with the total number of scientists and the average request per scientist growing faster than the pot. And as for publications, David Walker estimates that in the United States there are approximately 1500 journals that publish primary research *(2)*.

Who can keep up with the rate at which information is produced? In every field, senior people say they cannot stay on top of their specialties and junior people say there is no reward for learning anything outside a narrow area. The number of journals, reports, preprints, and conferences creates a demand for attention far in excess of the time one can devote to them. At the same time, it is very hard to locate information about projects related to one's own work. Competition among researchers seeking fame or fortune has bred secrecy rather than openness. Highly specialized subgroups form easily and take on lives of their own: a few hundred people with common interests can publish their own journal without outside contact. Don Swanson has even documented a case where two groups produced an extensive literature on the same medical phenomenon without knowledge of each other *(3)*.

The conclusion is inescapable: production of scientific information is exceeding anyone's ability to assimilate and use it. The body of scientific information is so far beyond the grasp of individuals and small groups that it is becoming ever more fragmented and disorganized. What can be done?

To support the dissemination and use of recorded knowledge, several computing technologies can help. Databases can manage information, bibliographies, access to facts and data, and cross referencing. Algorithms can package procedures in forms that can be used reliably by those who do not know their details. Expert systems can store and apply rules for processes of inquiry. Local area networks can link databases, algorithms, and expert systems with other processors such as simulators, statistical analyzers, and equation solvers. Concurrent processing can support "agents" that automatically seek out information while their owners are occupied with other tasks. And finally, heuristic searches can help sift through databases looking for new patterns and correlations.

But this use of computing technology is not enough. It deals only with packaged and recorded knowledge, ignoring the information and expertise that live in people. A new paradigm is needed to support the entire science knowledge base.

The new paradigm will place people in a central role, supporting their conversations and collaborations. It will use networks of computers to nurture networks of people. Here the term network refers not merely to a data channel, but to all the individuals, institutions, and services that must communicate in pursuit of science. The network augments communication and sharing; it is both a learning device for novices and a tool kit for experts.

Computer networks that nurture networks of people are the culmination of a process of evolution which can be said to have five stages:

1. file transfer

2. remote connections

3. distributed compution

4. real time collaboration

5. coherent function

At the first stage, the network is able to transfer files of information among computers, but without guaranteeing delivery time; this stage is sufficient to support electronic mail, bulletin boards, news services, and jointly authored papers. At the second stage, the network enables a user to connect to remote resources, such as instruments, computers, or databases, and employ them in real time as if they were local. At the third stage, the network is able to support distributed computations that include computing processes and resources at widely separated nodes; an example is a user interface process on a workstation, connected to a numerical process on a supercomputer, connected in turn to a graphics display system. At the fourth stage, the network directly supports collaboration by permitting real time conferences of users at different workstations, who can communicate as if they were gathered around one workstation -- that is, they can tap into a "common universe" in which they can talk, point to and share objects, edit and run programs, and examine outputs. At the fifth stage, the network is a coherent system comprising people and the resources contributed by them; each person can look in at this world from his workstation. The network will provide services to help people locate, use, and contribute resources, and to translate between the terminologies of the disciplines.

Most personal computer networks today are at the file transfer stage, offering only primitive remote connections. The ARPANET operates at the remote-connection stage and supports prototype subsystems for distributed computing and real time conferences *(4)*. No existing network is in the final stage of evolution, but government agencies are struggling at the beginning of it -- an internet, a loose federation of community networks sharing common protocols and gateways. The internet establishes a cooperation between communities, not a new centralized institution *(5)*.

The evolution of networking toward a worldwide system will facilitate the transition to the new paradigm of science. Conversations, collaborations, and professional relationships will spring up among participants around the world. Search agents will scan knowledge bases around the network looking for common patterns or new patterns, and report their findings to researchers. New ways to disseminate scientific knowledge will come into existence, the definition of publication will change, intellectual property rights will be sharpened, and there may even be means to evaluate submitted "papers" before releasing them for distribution. The new paradigm will indeed permit us to stand on one another's shoulders rather than step one another's toes.

Let me emphasize that the new paradigm is not merely a library function. Libraries are important repositories for recorded and packaged knowledge, but do not support the critical function of networking. They will participate in the new paradigm but will not dominate it.

When computing technologies support both the recorded and living parts of the knowledge base, scientists will have new means of building, creating, storing, accessing, exchanging, assimilating, integrating, and using knowledge and know-how. Besides serving science, these technologies will promote closer communication among all peoples, supporting everything from business transactions to professional relationships. And any technology that puts people into closer communication cannot help but be a stabilizing influence in a jittery world.

We stand at the threshold of a new era of science. The glut of information is pushing us over that threshold, and information processing technology is pulling us.

References

1. L. Lederman. 1987. "Science and technology policies and priorities: A comparative analysis." *Science 237*. 1125-1133.

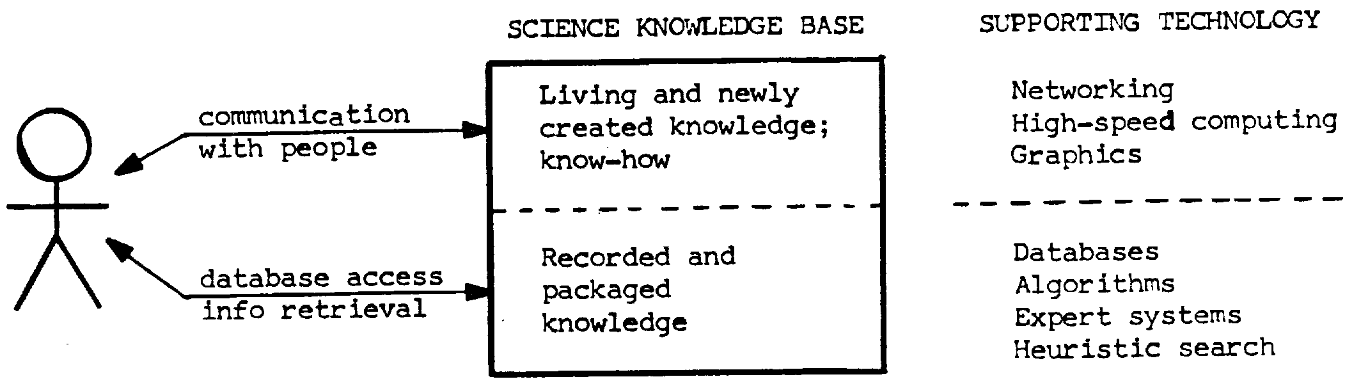
2. David A. Walker. 1987. "Management of scholarly publications." *Scholarly Publishing*. April. 189-196.

3. Don R. Swanson. 1987. "Two medical literatures that are logically but not bibliographically connected." *J. ASIS 38*, No. 4. 228-233.

4. Peter J. Denning. 1985. "Computer Networks." *American Scientist 73* No. 2 (March-April). 127-129.

5. Peter J. Denning. 1985. "Supernetworks." *American Scientist 73*, No. 3, May-June. 225-227.

BOX: Recording and Creating Knowledge



The science knowledge base consists of a body of information that is recorded and packaged for ease of retrieval and use, and a body of information constantly in the process of being brought forth in unpredictable ways from people's experience and knowledge. An individual gains access to the recorded and packaged part by accessing and retrieving information -- journals, libraries, or databases -- and to the creative part by communicating with the people in whom it lives. Technological support for the recorded part is based on mass storage and fast algorithms for retrieving and processing. Technological support for the creative part is based on networking interpreted in the broad sense of interchange and collaboration among people, and also on high-speed computation and graphics. The new paradigm for science will integrate computing technologies into the processes of scientific investigation. Indeed, these technologies are essential to surmount the growing glut of scientific information.

BOX: Requirements for Automating the Science Knowledge Base

Organizing large and changing sets of scientific information
into useful forms
Packaging knowledge in reusable forms
Sharing knowledge
Locating knowledge
Adding new knowledge efficiently to the knowledge base
Representing expertise, know-how, and skill
Certifying and validating information in the knowledge base
Correcting misinformation
Linking knowledge bases while respecting privacy and
confidentiality of personal information
Making networking available as ubitiquous service
Permitting collaborations at a distance and promoting
conversations among scientists
Using instruments remotely
Linking many resources into a single computation
Visualizing data